4.0

DRAFT

WHAT IS EPIDEMIOLOGY/DRAFT

Epidemiology is the branch of science that seeks the agents that cause disease in populations. Literally, it means the study of "things" (epi) which are "upon the people" (demos). Unlike the laboratory experiments we generally associate with science, epidemiologic studies are more like surveys. Through interviews and questionnaires, they gather information from groups of people that are similar in every way possible (age, sex, income level, geographic location, etc.) except for the one agent or exposure in question. These groups then are compared to determine a correlation between the exposure (asbestos, radon, ETS) and some effect (asbestosis, cancer, heart disease).

In the last 100 years, epidemiology has been important in tracing the causes of many diseases that might otherwise have gone undetected. In the 1800s, for example, people who contracted rabies were found to have one "exposure" in common. They had all been bitten by dogs. Researchers deduced that the infective agent was transmitted in dogs' saliva, thus leading to control of the disease.

More recently, epidemiology is credited with determining the method of transmitting the AIDS virus. Epidemiologists compared people infected with the disease to non-infected people who were

2503017328

Source: https://www.industrydocuments.ucsf.edu/docs/hgfk0000

DRAFT

otherwise similar to determine a common factor or exposure. On the surface, homosexuals, intravenous drug users and hemophiliacs were not groups one necessarily would link together. Through a process of elimination, epidemiologists concluded the common risk factor to be the exchange of body fluids.

Other epidemiological successes include identifying the cause of smallpox, the relationship between rubella in early pregnancy and birth defects, the dental effect of fluoridated water, and the relationship of dietary deficiency to pellagra.

Epidemiologic studies are increasingly popular today as most contemporary diseases are chronic, degenerative diseases that have several contributing factors, some of which have to do with some type of habit or behavior that theoretically can be changed. But each contributing cause may play only a small role, making it difficult to suggest a single cause for a specific effect. For example, high fat consumption has been identified as a risk factor for heart disease, but it is only one of several. It is difficult, if not impossible, to determine the degree of effect of one such factor. Thus, epidemiology at best can only <u>suggest</u> causal relationships, which must be confirmed some other way.

THE EPIDEMIOLOGIC STUDY

Despite its usefulness, epidemiology is inherently flawed as it relies on subjective rather than objective observation. Unlike

2503017329

Source: https://www.industrydocuments.ucsf.edu/docs/hgfk0000

DRAFT

other branches of science subject to stringent research standards, epidemiology has only limited safeguards against personal bias and numerous other variables that can interfere with a study's credibility.

The two principle types of epidemiologic studies are the "case-control" and "cohort" studies. The case-control study begins with people who already have the disease or exposure in question ("case" patients). They are compared to a similar group of people without the disease or exposure ("control" patients) to determine whether there is a correlation between the exposure and the disease. Thus, an epidemiologic study to determine whether radon exposure is a risk factor for cancer people with the disease who have been exposed (cases) to people who have not been exposed, but whose lifestyles and other environmental exposure were otherwise similar (controls).

Cohort studies start before people have a disease, but who are exposed to a suspected risk factor. Researchers follow these people over a long period of time and compare the frequency with which the disease in question occurs with the frequency it occurs in a group not exposed to the risk factor. Because of their higher degree of objectivity, cohort studies generally are more highly regarded than case-control studies. They are less popular among researchers, however, because of the difficulty in a transient society of monitoring large numbers of people over long periods of time.

2503017830

Source: https://www.industrydocuments.ucsf.edu/docs/hgfk0000

DRAFT

Case-control studies are most popular because they are easiest to conduct. They also are notoriously unreliable. Because they are conducted retroactively - after the fact of the disease and, oftentimes, after the patient's death, case-control studies are like trying to put together a puzzle without all the pieces. Researchers have to recreate the missing pieces by asking questions, often through questionnaire or telephone interviews, usually with friends or family of the deceased. If, for example, a woman dies of liver disease and the epidemiologist is trying to establish that the disease was related to alcohol consumption, he'd have to determine how much alcohol the woman drank.

The researcher might have to rely on the recollection of the woman's spouse or child. Few of us can accurately remember such details from our childhoods. Interviewers also can affect the reliability of the answers. A U.S. Census study found that if respondents know the purpose of the questionnaire, for example, they may try to give the "right" answers. Another study showed that people tend to put themselves (and possibly their mothers) in the best possible light. Men, for instance, typically exaggerate their height, and women significantly underestimate their weight, according to a New Zealand study. Hardly a criminal act, but personal bias can significantly alter the outcome of a scientific study dependent upon the accuracy of such answers.

The difficulties of conducting problem retroactive studies are

2503017331

Source: https://www.industrydocuments.ucsf.edu/docs/hgfk0000

DRAFT

compounded by the fact that subjects frequently are selected on the basis of death certificate information. Oftentimes, the death is attributed to the most obvious affliction based on the attending physician's knowledge of the patient. A death might erroneously be attributed to brain cancer when the brain cancer was, in fact, a metastasized cancer with the primary cancer occurring elsewhere. Sometimes, too, doctors who know a patient's lifestyle may search for a likely cause and, finding one, ignore other causes that might have contributed to the disease or death. Such misclassifications can skew the results of an epidemiologic study.

Even more important to the credibility of a questionnaire-based study is the number of people interviewed. The epidemiologic study resembles a marketing survey or political poll in that the larger the sample, the greater the accuracy. While most political polls use 1,000 as a minimum sample size, epidemiologic studies can be conducted on as few as 10 people, thus increasing the likelihood that the results are meaningless. Given that a study of 10 isn't a representative cross-section of a population, one incorrect response could significantly alter the conclusion. Making causal assumptions on the basis of such small studies, which occurs all too frequently, is both foolish and irresponsible.

CONFOUNDING FACTORS

Another flaw in epidemiologic studies is the near impossibility of finding groups of people who are alike in every

2503017332

Source: https://www.industrydocuments.ucsf.edu/docs/hgfk0000

DRAFT

way except for the exposure in question. While it's relatively easy to separate groups by age, sex and weight, it's more difficult to match groups based on diet and exercise, both independent factors that might be associated with the disease in question. It is, for now, impossible to match groups by genetic background, which many consider the strongest independent factor of all.

Moreover, certain types of behavior or exposures tend to occur together. For example, studies show that people who smoke also tend to drink more and pay less attention to nutrition than nonsmokers. In trying to determine whether smoking contributes to a particular disease, the epidemiologic study must determine whether the association is with smoking, drinking or another factor. These other possible influences are called "confounding factors" (as in Confound It!) and must be eliminated to make a scientific determination of association or increased risk. Clearly, it is difficult to eliminate all other possible influences. There are statistical methods of neutralizing the effect of confounding variables, but they're not foolproof. They're also useless if the confounding variables are unknown.

[redacted] THE STATISTICS [redacted]

Confounding factors, medical misclassification, personal bias and flawed recollections all contribute to epidemiology's limited usefulness. Beyond these are the intrinsic limitations to statistical analysis, which is used to interpret the epidemiologic

2503017333

Source: https://www.industrydocuments.ucsf.edu/docs/hgfk0000

DRAFT

data and to determine the relative risk of some exposure. Statistical analysis is integral to experiment and observation but, as Dr. William W. Lowrance of Harvard put it: "Statistics cannot decide for us what is truth and what is not. After all, these frail numbers tenuously imposed on the situation are ours, not Nature's."

Many of perceived risks today are, in fact, statistically contrived rather than real risks. These contrivances occur as a result of a practice known as "data-dredging," in which researchers compile groups and subgroups of data gathered during epidemiologic studies to produce a variety of associations. With modern electronic computation, diverse information can be assembled relatively easily. These might include demographic data (age, race, sex, socioeconomic status), data about exposures and lifestyle (diet, smoking, alcohol, environmental agents), and data about disease and other health effects (birth defects, stroke, heart disease, cancer, death).

In data dredging, some or all of these data may be combined statistically to produce various associations between exposures and diseases within certain demographic groups. If any of these combinations deliver an association that can be considered "statistically significant, then a cause-effect relationship can be proposed. The problem with such statistical conjecture is that the effect may be due to chance or, in the absence of adequate controls for bias, misclassification and confounders, merely a statistical

2503017334

Source: https://www.industrydocuments.ucsf.edu/docs/hgfk0000

DRAFT

coincidence.

One such instance occurred a few years ago when epidemiologists pronounced an association between alcohol use and breast cancer. This claim was based on a review of questionnaires filled out over several years by 122,000 nurses. The data had been gathered for some other purpose than to determine a link between alcohol and breast cancer. Through statistical analysis, the epidemiologists determined that nurses who reported regular alcohol consumption were 60 percent more likely to report development of breast cancer than non-drinking nurses. The association was formulated on the basis of a statistical relationship, rather than as a result of a study specifically designed to explore the association.

There are numerous problems with this sort of retrospective hypothesizing. In a review of the alcohol-breast cancer study, Dr. Alvan R. Feinstein, Harvard professor of medicine and epidemiology, criticized the researchers for making erroneous claims on the basis of such spurious associations. He pointed out that the researchers themselves had not examined the nurses for breast cancer and, thus, didn't know how many undiagnosed cases there might have been among them. They relied only on the nurses' reporting of the cancer, a well as their own recollection of alcohol consumption - "a substance whose recorded intake is notoriously inaccurate," Feinstein said.

2503017335

Source: https://www.industrydocuments.ucsf.edu/docs/hgfk0000